ELSEVIER

Biochimica et Biophysica Acta 1339 (1997) 39–52

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica*

Li Ye [1], Lys M. Buck [2], Holly J. Schaeffer [3], Franklin R. Leach *

Department of Biochemistry and Molecular Biology, Oklahoma State University, 246 NRC, Stillwater, OK 74078-3035, USA

Received 22 May 1996; revised 4 November 1996; accepted 5 November 1996

Abstract

The first cDNA from the Photurinae subfamily of the Lampyridae encoding a firefly luciferase from lantern mRNA of *Photuris pennsylvanica* has been cloned, sequenced, the amino-acid sequence predicted and the sequence reported to GenBank. The cDNA was about 1.8 kb in length with the largest open reading frame coding for a 545-residue protein. The 5′ noncoding region is 61 bp long and the 3′ noncoding region is 135 bp in length. There is a 24-nucleotide poly(A) tail. When the amino-acid residues are aligned, *P. pennsylvanica* contains 154 (about 28% of the total residues) that are conserved in all 16 of the deduced luciferase sequences that are presently available. In this *P. pennsylvanica* luciferase, the amino acids at 276 of the positions are the same at corresponding positions of at least one of the other enzymes. There are two amino-acid differences between this luciferase and the unpublished sequence obtained by Dr. Keith Wood for a putative larval *Photuris* firefly luciferase cloned from a Maryland firefly. Signature amino-acid sequences and domains found in the deduced sequence are for adenylate kinase, the putative AMP-binding domain, luciferin 4-monooxygenase, 4-coumarate CoA ligase, long-chain fatty acid CoA ligase, 2-acylglycerophosphoethanolamine acyltransferase, the microbody-directing sequence, peptide-synthesizing complexes, and acyladenylate-synthesizing enzymes.

Keywords: Luciferase; Bioluminescence; Luciferin; Primary structure; (*P. pennsylvanica*); (Firefly)

1. Introduction

Firefly luciferase (*Photinus* luciferin:oxygen 4-oxidoreductase, EC 1.13.12.7, abbreviated LUC) produces light by the oxidative decarboxylation of luciferin (LH_2) as shown in the following equations that represent the two-step reaction.

Step one forms an enzyme-bound luciferyl adenylate

$$LH_2 + MgATP + LUC \rightleftharpoons LUC\text{-}LH_2\text{-}AMP + MgPP_i$$

Step two is the oxidative decarboxylation of luciferin with the production of light upon decay of the excited form of oxyluciferin

$$LUC\text{-}LH_2\text{-}AMP + O_2 + OH^-$$
$$\rightarrow LUC\text{-}OL + CO_2 + AMP + light + H_2O$$

There is a slow release of the oxyluciferin product, OL, from the enzyme-product complex.

The enzymatic reaction has a quantum yield of 0.88 photon/molecule of luciferin oxidized [1]. The

* Corresponding author. Fax: +1 (405) 7447799; E-mail: firefly@biochem.okstate.edu

[1] Present address: Department of Molecular Biology, Holland Laboratory, American Red Cross, Rockville, MD, 20855.

[2] Present address: 1525 SW 7, #8, Albany, OR 97321.

[3] Present address: Department of Horticulture, Michigan State University, East Lansing, MI 48824.

PII S0167-4838(96)00211-7

Table 1
Cloned firefly luciferases

Year	Species	Abbrev. LUC	GenBank	sp	PIR	Locus	% same	# Amino acids	λ_{max}	Reference
1987	*Photinus pyralis*	*Ppy*	M15077	P08659	A26772	PPYLUC	60.4	550	562	[7]
1989	*Pyrophorus plagiophthalamus* (4)									[8]
	Green-LucGR	*Ppl*GR			S29352		48.0	543	546	
	Yellow green-LucYG	*Ppl*YG			S29353		48.9	543	560	
	Yellow-LucYE	*Ppl*YE			S29354		49.3	543	578	
	Orange-LucOR	*Ppl*OR			S29355		49.1	542	593	
1989	*Luciola cruciata*	*Lcr*	M26194	P13129	JS0181	FFLLUC	55.3	548	562	[9]
1992	*Luciola lateralis*	*Lla*	X66919	Q01158	S23437	LLLUCI	55.5	548	552	[10]
1993	*Luciola mingrelica*	*Lmi*	S61961		S33788	S61961	53.7	548	570	[11]
1995	*Hotaria parvula*	*Hpa*	L39929			HOTLUCI	55.3	548	568	[12]
1995	*Pyrocoelia miyako*	*Pma*	L39928			PIBLUCIF	60.1	548	550	[12]
1995	*Lampyris noctiluca*	*Lno*	X89479			LNLUCPROT	61.7	547	550	Gene bank, [13]
1995	*Luciola lateralis* gene	*Lla*(g)	Z49891			LLLUCIFMJ	56.2	548	552	Gene bank
1994	*Photuris pennsylvanica*	*Ppe*(LY)	U31240			PPU31240	–	545		This lab, [14]
1994	*Photuris pennsylvanica*	*Ppe*J19					58	552		Patent *
	Photuris pennsylvanica	*Ppe1*(KW)					55	552	560	Wood **
	Photuris pennsylvanica	*Ppe2*(KW)					100	545	538	Wood **

From GenBank, Swiss Protein Sequence Data Base at ExPASy (sp), and Protein Identification Resource at Johns Hopkins (PIR).
* CAS-registry 160831-30-1 protein sequence for JP 94303982 A2, not entered into Chemical Abstracts until after Li Ye thesis was submitted [14]; ** Keith Wood, personal communication. The abbreviations for the LUC for specific organisms are adapted from those used by Wood (personal communication). Where both a cDNA and a gene are deposited, the gene sequences is indicated by a (g) after the standard abbreviation. When the same species has two or more sequences published, a Roman numeral is added. For sequences obtained in different laboratories the initial of the principal investigator are added in parentheses. When luciferase from one species produce more than one color light, a two letter abbreviation code suggested by Wood et al. [8] is used. The % same column shows the percentage of identity with *P. pennsylvanica* sequence. The wavelength of light produced is based on that by the beetle unless more than one species exist and then it is based on that determined using the enzyme.

enzyme is widely used in the quantitation of ATP and other biochemically important compounds [2–5] and as reporter of gene expression (see bibliographic lists in J. Biolumin. Chemilumin. 5, 141–152 (1990) and 8, 267–291 (1993)).

The gene for firefly luciferase of the North American firefly, *Photinus pyralis*, was cloned and sequenced by DeLuca and colleagues [6,7]. Since those studies, firefly luciferase cDNAs or genes have now been cloned from several other beetle species; these are listed in Table 1. This paper reports the cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica*; this is the first species from the Photurinae subfamily to have its luciferase sequence determined and reported to GenBank. *Photuris pennsylvanica* is a twilight/night-active firefly while the common, well-characterized North American species, *P. pyralis*, is dusk-active, flashing only during twilight. Before mating, *Photuris* fireflies females respond to courtship signals of conspecific males [15]. After mating, they become '*femmes fatales*' by answering the courtship flashes of males of other species who are then treated as prey. Predation by aggressive mimicry is known only for *Photuris* [16].

The *Photuris* genus has been mainly studied from the biological standpoint. The biochemical and structural changes that occur during light organ development have been studied by Strause et al. [17]. During development the larval light organ regresses and is replaced by the adult lantern. During pupation the levels of luciferase and luciferin remain constant in the posterior half of the pupa while there is an initial increase followed by a decrease of luciferase and luciferin in the anterior half. Strause and DeLuca [18] found a luciferase isozyme in larval *Photuris pennsylvanica* that is distinct from the enzyme of the adult. This laboratory has identified two firefly luciferases from adult *P. pennsylvanica* lanterns during Sephadex G-150 chromatography [14].

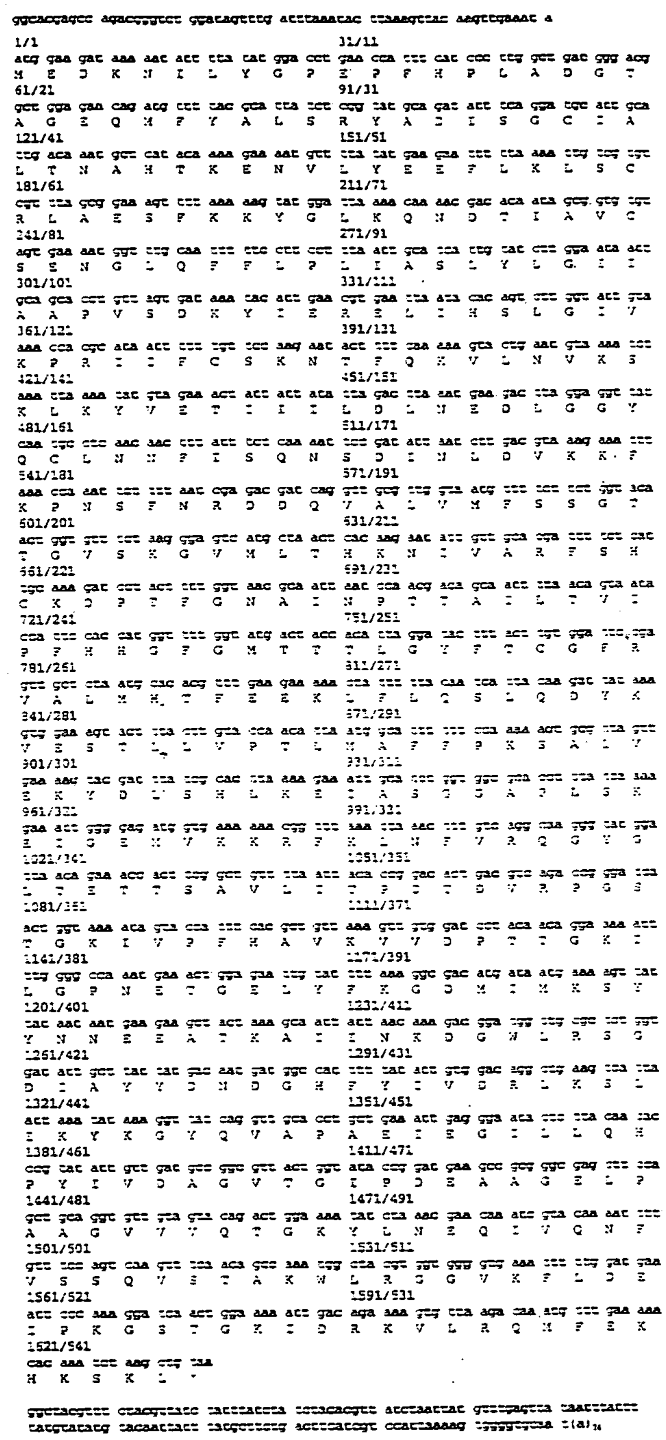


Fig. 1. Nucleotide and predicted amino acids sequences for cDNA from *Photuris pennsylvanica*. The 61 bp leader is shown in groups of 10 bp with a residual. The coding sequence starts at bp 62 and continues through bp 1700. The 5′ and 3′ noncoding region is shown in groups of 10 bp. The figure was prepared by editing a DNA Strider report after the consensus sequence was derived using the AssemblyLIGN program.

The overall goal of this study is to increase knowledge concerning amino acids that might function in catalyzing light production by luciferin oxidation by obtaining an unique sequence from another subfamily of fireflies. Comparison of the amino-acid sequences of several firefly luciferases may indicate which amino acids are functionally important since these should be conserved among the various species. The sequence of *P. pennsylvanica* luciferase has been deposited in the GenBank as entry U31240.

2. Materials and methods

2.1. Bacterial strains and media

E. coli strain XL1-Blue MRF′ (Stratagene Cat. # 200301) was used for λZAP Uni-Zap™ library plating. *E. coli* strain XL1-Blue (Stratagene Cat. # 200268) was used for amplification of λZAP Uni-Zap™ library and for recombinant plasmid replication and preparation. *E. coli* strain SOLR™ (Stratagene Cat. # 200298) was used as a plasmid host for in vivo excision and cDNA library screening.

2.2. Firefly collection

Fireflies were collected locally with nets and stored in jars until the evening's collection was completed. The live fireflies were taken to the laboratory, sorted according to species, and immediately frozen in liquid nitrogen. The frozen fireflies were stored in a −80°C freezer.

2.3. mRNA isolation and library preparation

The lanterns of the frozen *P. pennsylvanica* fireflies were removed under liquid nitrogen and then one g of lanterns was pulverized under liquid nitrogen with a mortar and pestle. The lantern RNA was isolated by using a Stratagene RNA Isolation Kit (Cat. # 200345, Lot 138) and then the mRNA was isolated using Stratagene's Quik™ mRNA Isolation Kit (Cat. # 200349, lot 133). A Stratagene ZAP-cDNA™ Synthesis Kit (Cat. # 200400, Lot UC122) was used for construction of the cDNA library using 5 μg mRNA. In vivo excision of pBluescript® phagemid from Uni-ZAP™ XR (Stratagene ZAP-cDNA™ Synthesis Kit) was done according to the instructions provided to produce a phagemid.

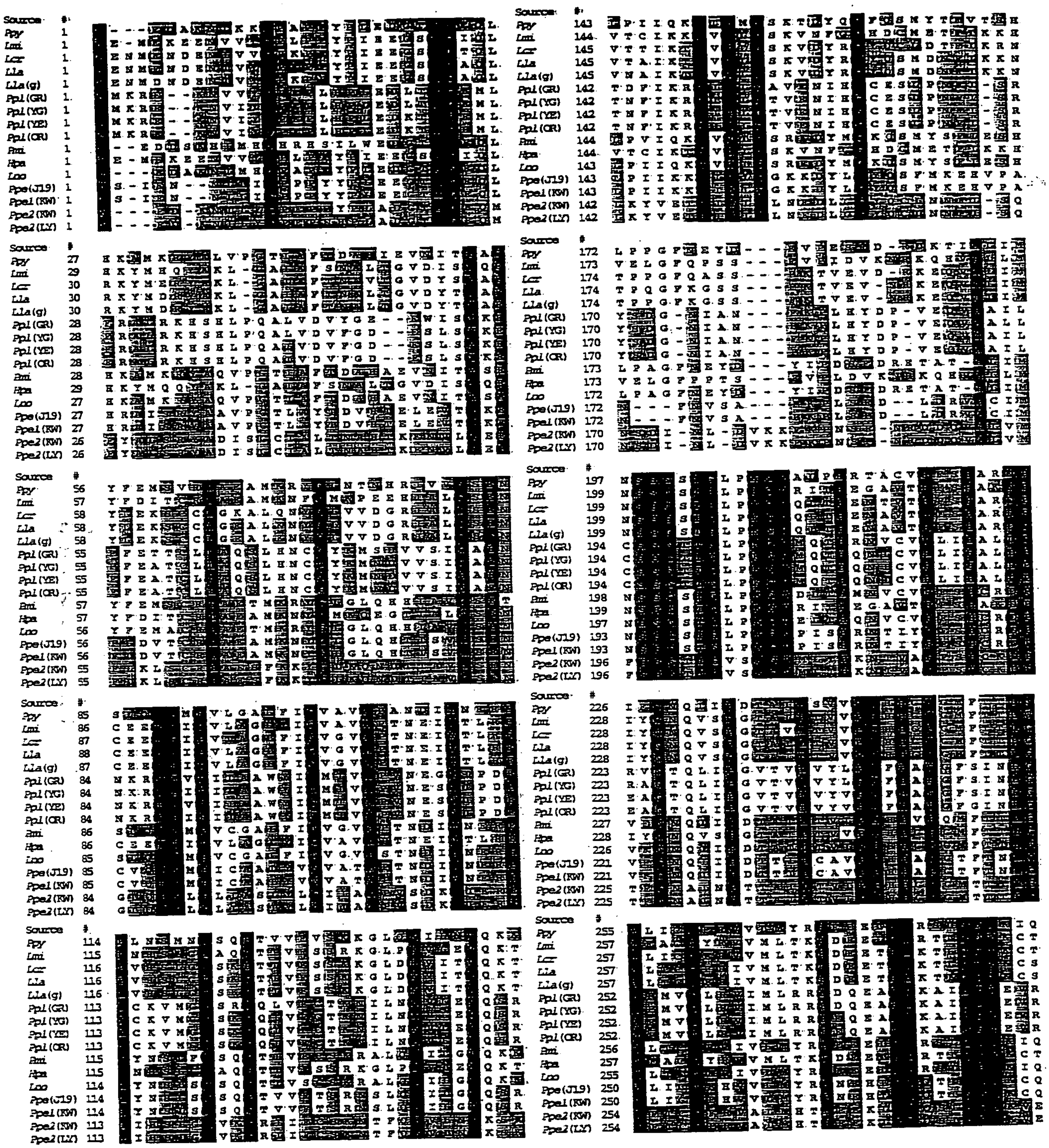
2.4. Expression

Expression was achieved by following a procedure similar to that of Devine et al. [11]. The instructions supplied with Stratagene's ExAssist™ kit were followed for excision. Twelve 150 mm plates were used for the first screening and incubated at 37°C overnight. The colonies on each individual plate were lifted onto a nitrocellulose filter and the filters were placed on fresh LB-ampicillin plates. The master plates and lifted filters were incubated at 37°C for another 2 h. The filters with colonies were switched to 22°C and incubated for another 2 h for expression. The filters were removed from the plates, soaked with 1 mM firefly luciferin (in 0.1 M Tris acetate, 10 mM $MgSO_4$, 2 mM EDTA, pH 5.0) for 5 min, wrapped with Saran Wrap® and exposed to X-ray film overnight. After film development, the positive colonies (bioluminescent) were identified (they also could be observed by eye after dark-adaptation). Two colonies were picked from the master plates, and streaked onto fresh LB-ampicillin plates. Positive colonies were purified by two more rounds of plating and screening.

2.5. Sequencing

The pBluescript® phagemid was used for sequencing. DNA sequencing was done by the Sarkeys Biotechnology Laboratory in this Department. The sequencing was started using T3 and T7 primers.

Fig. 2. Comparison of the amino-acid sequences from sixteen firefly luciferases. An alignment of the amino-acid residues from the sixteen firefly luciferase listed in Table 1 was prepared. The shading used indicate: ■ an amino-acid residue conserved in all 16 luciferases; ▤ a residue found in *Photuris pennsylvanica* and at least one other firefly luciferase sequence; and –, a gap in the sequence used for alignment of the residues. The alignment was handmade and checked by comparison to an alignment made using the MACAW, and Clustal programs.



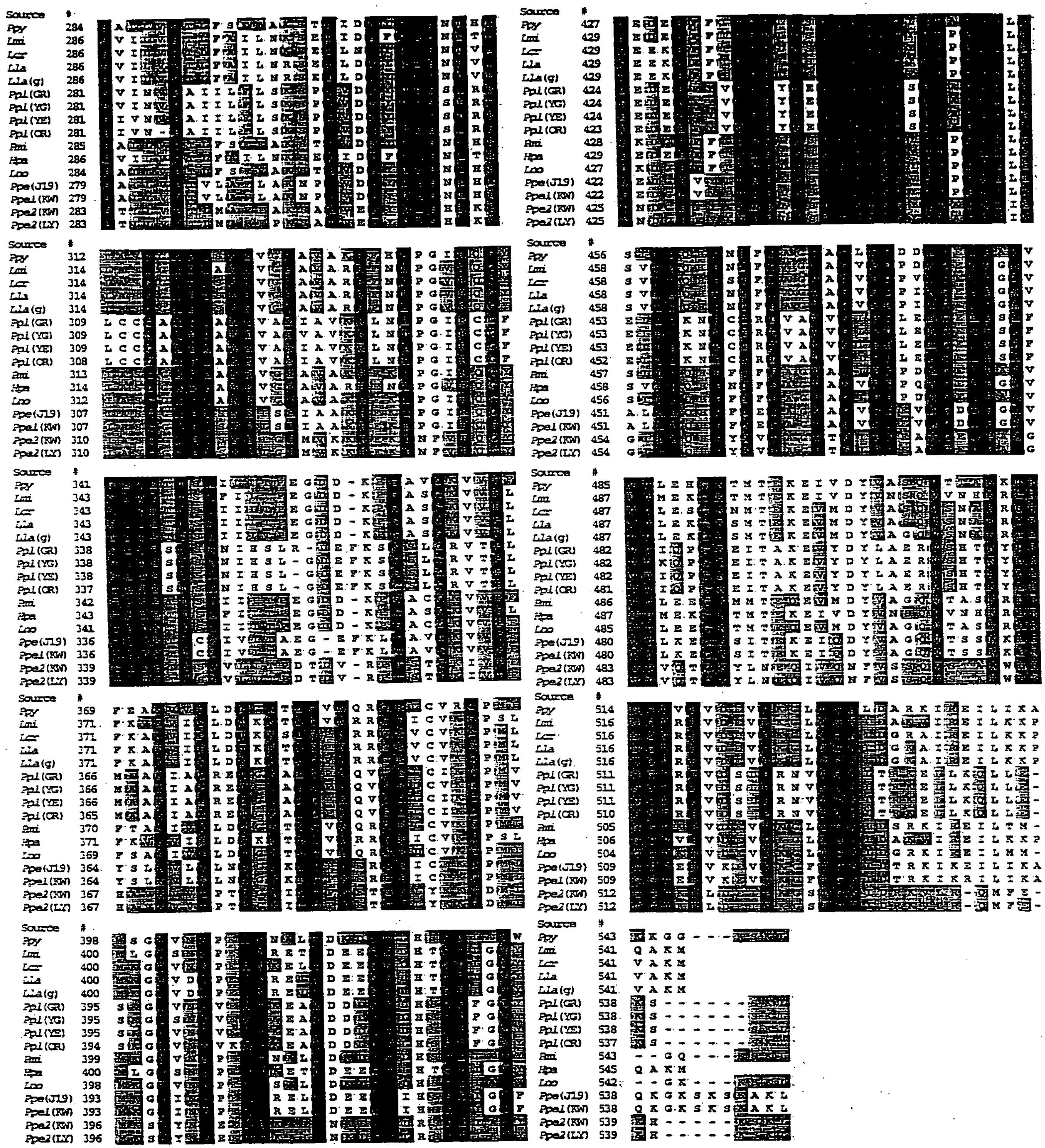


Fig. 2 (continued).

Subsequent sequencing was continued in both directions using primers designed using the Oligo® 4.06 program (National Biosciences, Inc.) based on the determined sequence. The primers were synthesized by the Sarkeys Biotechnology Laboratory.

2.6. *Measurement of luciferase gene size*

The plasmid containing the luciferase gene was digested with *Eco*R I and *Xho*I at 37°C overnight. A separate digestion was done using *Bam*HI and *Kpn*I under the same conditions. The digested samples were separated by 1% agarose gel electrophoresis with a 1 kb DNA ladder (BRL, Cat. No. 1561057) as a standard and uncut plasmid as a control.

2.7. *Computer analysis of data*

The following programs and data bases (versions) were used: AssemblyLIGN, V 1.0.7; Beauty [19], BLASTPAT, and various other search algorithms via the BCM Search Launcher of the Human Genome Center at the Baylor College of Medicine, Houston TX; Blocks [20,21]; DNA Strider, V1.2 [22]; GenBank, NCBI, release 89.0 [23]; MacVector, V 4.5.2 [24]; p*I*/MW calculated at the ExPASy Server; ProDom, release 28 [25,26], ProSite [27]; PROSITE, release 13 [27]; and SWISS-PROT, release 21.0 [28].

3. Results

3.1. *Isolation and sequencing of cDNA*

The cDNA library was prepared from the lanterns of locally collected *P. pennsylvanica* fireflies, expressed in *E. coli*, and screened for light production after luciferin addition. Since the screening detected expressed bioluminescence, only functional cDNA sequences were identified. The insert size as determined after *Eco*RI and *Xho*I digestion and 1% agarose gel electrophoresis was about 1.8 kb which is sufficient to code for the entire luciferase polypeptide of approx. 550 amino-acid residues (based on the length of other firefly luciferases). In a restriction enzyme-based analysis, the selected clone did not contain a *Kpn*I restriction site but did have a single *Bam*HI site within the luciferase-coding sequence (which is not found in the *Photinus* sequence). No *Kpn*I restriction site is present in the *Photinus* sequence. Devine et al. [11] found clones of *Luciola mingrelica* cDNA that had either one or two *Kpn*I sites.

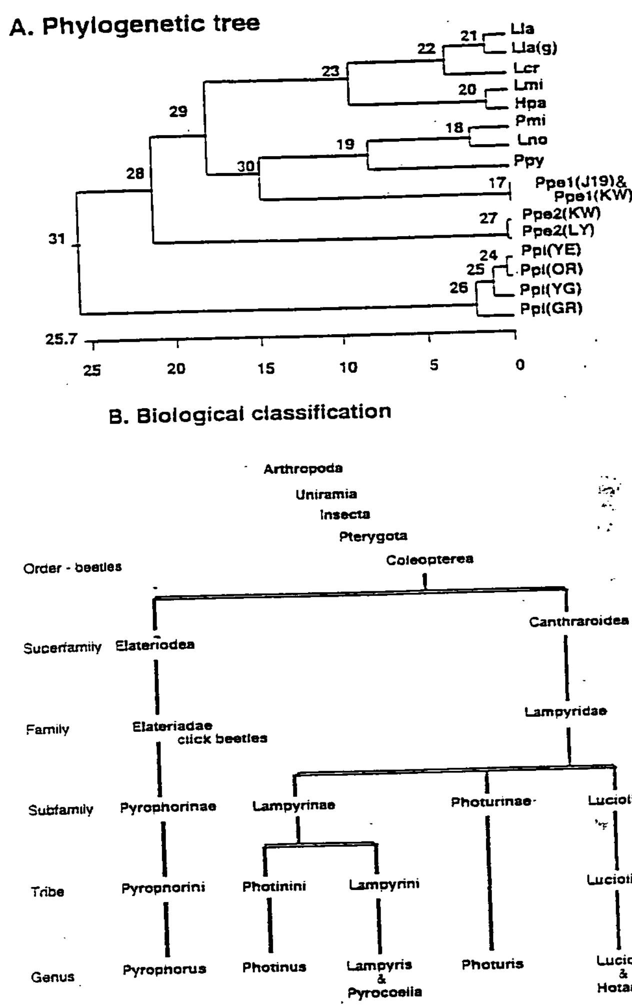


Fig. 3. Relationships among the amino-acid residues and the biological classification of the sixteen sequenced firefly luciferases. (A) The relationship among firefly species based on firefly luciferase amino acid sequences as determined using the protein parsimony and the Clustal algorithm in the DNA Star program. The length of each pair represents the distance between the sequence pairs and the scale beneath the tree measures the genetic distance between sequences. (B) Biological classification of the firefly species from which the luciferase has been sequenced. This classification of firefly species was adapted from Herring [29] and Campbell [5].

Table 2
Amino-acid occurrences in sixteen firefly luciferases

	Firefly luciferase source															
	Luc															
	Ppy	Lcr	Lla	Lla(g)	Lmi	PplGR	PplYE	PplOR	PplYG	Hpa	Pmi	Lno	Ppe1(J19)	Ppe1(KW)	Ppe2(KW)	Ppe(LY)
MW	60706	59979	60087	60010	60456	60652	60412	60296	60268	60327	60917	60365	60931	60931	60610	60610
p*I*	6.42	7.07	6.50	7.96	6.24	6.69	6.39	6.69	6.69	6.27	6.12	6.09	7.55	7.55	8.39	8.40
Amino acid																
Total	550	548	548	548	548	543	543	542	542	548	548	547	552	552	545	545
Nonpolar																
A	42	32	34	34	33	36	37	37	37	33	36	39	36	36	34	33
C	4	8	7	7	8	13	13	13	13	8	9	9	11	11	7	7
F	30	23	23	23	27	23	24	24	24	26	29	28	25	25	30	30
I	38	33	36	36	36	39	42	43	38	35	39	39	42	42	39	39
L	52	49	49	49	48	57	54	54	56	49	49	49	54	54	54	54
M	14	11	12	12	13	11	11	11	11	13	18	17	9	9	11	11
P	29	29	28	29	30	28	28	28	28	30	26	28	30	30	22	23
V	44	55	49	49	50	49	47	45	48	50	39	37	41	41	38	38
sum	253	240	238	239	245	256	256	255	255	244	245	246	248	248	235	235
No charge																
G	45	51	52	53	49	38	38	39	39	50	46	49	46	46	43	43
N	19	20	20	20	19	21	23	23	23	19	13	14	21	21	25	25
Q	16	13	12	11	14	13	14	14	14	16	20	19	18	18	17	17
S	29	30	29	28	26	31	32	31	31	25	21	32	32	32	33	33
T	29	36	34	34	32	19	19	19	19	35	31	29	29	29	34	34
W	2	1	1	1	1	3	2	2	2	1	2	1	1	1	2	2
Y	19	21	22	22	19	21	20	20	20	20	21	22	24	24	24	23
sum	159	172	170	169	160	146	148	148	148	196	154	166	171	171	178	177
Charged																
D	31	25	27	28	25	26	25	25	26	24	32	30	23	23	27	27
E	33	39	40	37	42	39	40	39	38	41	33	34	36	36	31	31
H	14	8	8	8	15	13	13	13	13	13	17	14	14	14	12	13
K	40	43	44	46	42	35	37	38	37	42	36	37	42	42	47	47
R	20	21	21	21	19	28	24	24	25	18	31	20	18	18	15	15
sum	138	136	140	140	143	141	139	139	139	138	149	135	133	133	132	133

The amino-acid composition and molecular weights were calculated using the DNA Strider program with input data from GenBank, patent information, personal communications, and PIR. p*I* was calculated by the ExPASy p*I*/MW program.

Table 3
Percentage of idenity and similarly among pairs of the 16 firefly luciferases

	PPy	*Lmi*	*Lcr*	*Lla*	*Lla*(g)	*Ppl*(GR)	*Ppl*(YG)	*Ppl*(YE)	*Ppl*(OR)	*Pmi*	*Hpa*	*Lno*	*Ppe1*(J19)	*Ppe1*(KW)	*Ppe2*(KW)	*Ppe2*(LY)
Ppy		67.9 (82.1)	60.4 (82.1)	68.6 (82.8)	68.2 (82.8)	48.7 (67.5)	48.6 (67.6)	48.7 (67.7)	48.4 (67.7)	82.4 (88.8)	67.7 (81.7)	84.5 (90.7)	70.1 (83.6)	70.1 (83.6)	70.1 (76.2)	60.0 (76.0)
Lmi			80.8 (90.1)	82.1 (90.5)	81.5 (90.3)	48.4 (66.6)	48.9 (67.1)	47.9 (67.3)	48.0 (67.1)	64.7 (79.6)	98.0 (99.1)	66.2 (81.5)	61.6 (79.4)	61.6 (79.4)	53.8 (73.8)	53.6 (73.7)
Lcr				93.6 (96.9)	93.4 (97.1)	49.1 (67.5)	49.9 (67.8)	48.9 (68.1)	49.0 (67.8)	66.6 (79.8)	81.1 (89.9)	67.7 (80.9)	62.3 (80.3)	62.3 (80.3)	56.0 (73.9)	55.8 (73.7)
Lla					99.1 (99.5)	48.2 (67.3)	49.0 (67.8)	48.0 (68.1)	48.1 (67.8)	67.3 (80.6)	82.2 (90.5)	68.4 (81.8)	62.5 (80.3)	62.5 (80.3)	56.7 (74.8)	56.6 (74.6)
Lla(g)						47.8 (67.2)	48.6 (67.7)	47.6 (67.9)	47.7 (67.7)	67.1 (80.5)	81.7 (90.3)	68.3 (81.8)	62.5 (80.3)	62.5 (80.3)	56.4 (74.6)	56.2 (69.1)
Ppl(GR)							96.1 (98.5)	95.8 (98.3)	95.2 (98.2)	47.9 (66.5)	48.3 (66.2)	49.3 (68.1)	50.7 (69.9)	50.7 (69.9)	48.5 (68.9)	48.7 (69.1)
Ppl(YG)								97.8 (99.5)	98.0 (99.3)	47.6 (66.6)	48.8 (66.7)	49.2 (68.4)	50.3 (69.5)	50.3 (69.5)	48.5 (68.8)	49.4 (69.0)
Ppl(YE)									99.5 (99.8)	47.7 (66.7)	47.8 (67.0)	49.3 (68.5)	50.6 (69.7)	50.6 (50.6)	49.3 (69.1)	49.4 (69.3)
Ppl(OR)										47.4 (66.6)	47.9 (66.7)	49.0 (68.4)	50.3 (69.6)	50.3 (69.6)	49.0 (69.0)	49.2 (69.2)
Pmi											64.3 (79.2)	95.2 (96.3)	69.7 (82.8)	69.7 (82.8)	60.68 (75.8)	60.4 (75.6)
Hpa												65.9 (81.1)	61.6 (79.2)	61.6 (79.2)	54.7 (74.2)	54.5 (74.0)
Lno													71.0 (84.2)	71.0 (84.2)	61.8 (76.7)	61.5 (76.5)
Ppe1(J19)														100 (100)	57.5 (75.8)	57.1 (75.6)
Ppe1(KW)															57.1 (75.6)	57.5 (75.6)
Ppe2(KW)																99.6 (99.8)
Ppe2(LY)																

Calculated using the GCG program BestFit. Shown in parentheses are the calculated similarities.

Fig. 1 shows the nucleotide sequence of the *P. pennsylvanica* cDNA and the deduced amino-acid sequence for the largest open reading frame. From the sequence analysis, the cDNA is 1831 bp long with an open reading frame (ORF) of 1635 bp. The ORF encodes a protein of 545 amino acids with a calculated molecular weight of 60 610 daltons. The 5′ untranslated region contains 61 bp and the 3′ untranslated region has 135 bp. The 3′ noncoding region contains a poly(A) tail of 24 nucleotides.

3.2. Comparison of the deduced amino-acid sequences

The amino-acid sequences deduced from the 16 cDNAs and genes sequenced for firefly luciferases have been aligned (see Fig. 2) to allow determination of conserved amino-acid residues and suggest possible functional portions. There are 154 residues conserved among all the luciferases (about 28% of the total residues). In the putative *P. pennsylvanica* larval luciferase, the amino acids at 276 positions are the same at corresponding positions of at least one other species. One hundred and fifteen amino-acid residues are unique to the putative *P. pennsylvanica* larval enzymes. Of these, 24 residues are conserved in all other species.

The amino-acid compositions of the sequenced firefly luciferases are compared in Table 2. The calculated isoelectric points of putative *P. pennsylvanica* larval enzymes are the highest of all isozymes.

3.3. Relatedness

The amino-acid sequences of the firefly luciferases were analyzed by protein parsimony using the DNA Star program (Fig. 3A). The phylogenetic classification for these beetles is shown in Fig. 3B [5,29]. As expected, the various *Luciola* species are closely related. The relationships among the various species as determined by luciferase amino-acid sequence appear similar to the relationships based on biological classification. The percentage of identity and similarity were calculated for the 16 firefly luciferase sequences using the GCG program BestFit with a gap weight of 3.0 (Table 3). The *Ppe*1 and *Ppe*2 luciferases are 57% identical.

4. Discussion

4.1. Related sequences

When either the cDNA sequence or the predicted amino-acid sequence was used as the input sequence for computer-based searches for similarity, the high-scoring related sequences were the luciferases from both the Lampyridae and the Elateridae families, 4-coumarate CoA ligase, long-chain CoA ligase, 2-acylglycerophosphoethanolamine acyltransferase, and the peptide-antibiotic-synthesizing enzymes such as gramicidin S synthetase and tyrocidine synthetase. These relationships have been reported [30–38] for the other firefly luciferases.

4.2. Domain structure

A domain structure map for the predicted amino-acid sequence of *Photuris* firefly luciferase was developed by using the ProSite, ProDom, PRINTS, BLOCKS, and PepPepSearch programs. Fig. 4 illustrates these results and Table 4 defines the sites and their presumed functions, if known.

The T-250TLGYFT-256 sequence (see also Fig. 2) is the AMP binding block BL00455B. PS00339, the AA tRNA ligase II.2 sequence whose consensus sequence is [GSTALVF]-{DENQHRKP}-[GSTA]-[LIVMF]-[DE]-R-[LIVMF]-x-[LIVMSTAG]-[LIVMFY], was found as F-431YIVDRLKSL-440 (correct in 9/10 positions). The ProSite convention is [ambiguities where indicated amino acids are acceptable] and {amino acids not accepted in this position}. The G-338YGLTRYSAVLITPDTDVRPGSTG-362 sequence is the domain II that is conserved in acyl-adenylate-synthesizing enzymes [30]. The adenylate kinase signature, PS00113, with consensus sequence [LIVMFYW](3)-D-G-[FY]-P-R-x(3)-[NQ] was tentatively found as I-411NKDGWLRSGDI-422 (correct in 7/12 positions). The G-415WLRSGD-421 sequence is the domain III that is conserved in acyl-adenylate-synthesizing enzymes [30]. The SKL sequence at the C-terminus is the microbody-directing sequence and was detected by the SORT program (it is also ProSite PS00342).

The putative AMP-binding domain signature of [LIVMFV]-x(2)-[STG](2)-G-[ST]-[STE]-[SG]-x-[PALIVM]-K (ProSite P00455) was found as V-194MFSSGTTGVSK-205 and is highly conserved

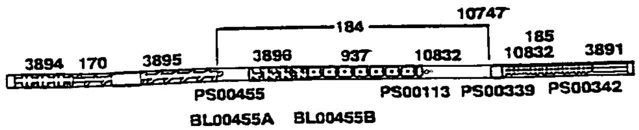


Fig. 4. Putative functional domains of *Photuris pennsylvanica* firefly luciferase. The figure is a composite of the BEAUTY map and ProDom, ProSite and Blocks maps of possible functional domains. The ProDom sequences are: 3894, luciferin 4-monooxygenase (firefly luciferase); 170, 4-coumarate CoA ligase; 3895, luciferin 4-monooxygenase; 184, gramicidin S synthetase; 3896, luciferin 4-monooxygenase; 937, 4-coumarate CoA ligase; 10747, 4-coumarate CoA ligase; 10748, 4-coumarate CoA ligase; 10832, 2acylglycerophosphoethanolamine acyltransferase; 185, gramicidin S synthetase; and 3891, luciferin 4-monooxygenase. The ProSite sequences are: PS00455, putative AMP-binding domain sequence; PS00113, adenylate kinase signature (similar); PS00339, aminoacyl-tRNA synthetase class II.2 (similar); and PS00342, microbodies C-terminal targeting sequence. The Blocks are: BL00455A and BL00455B, the putative AMP-binding domain.

among the various firefly species — the consensus sequence for luciferase is $I_{(11)}M_{(12)}N_{(10)}S_{(16)}S_{(16)}$-$G_{(16)}S_{(10)}T_{(16)}G_{(16)}L_{(14)}P_{(14)}K_{(16)}$ (where the subscript number in parentheses is the number of occurrences in the 16 sequences).

Among the sequences not found (whose possible existence was considered on the basis of similarity in function) is the consensus pattern of [FYH]-R-X-[DE]-X(4, 12)-[RH]-X(3)-F-X(3)-[DE], ProSite PS00179 which is the amino-acid tRNA ligase II 1 pattern. ProSites that were sought but not found include: the ATP/GTP-binding site motif A (P-loop) PS00017, the chloramphenicol acetyltransferase active-site PS0100, the protein kinase signatures (PS00107, PS00108, and PS00109), the ubiquitin-activating enzyme signatures (PS00536 and PS00865), and the acyl-CoA-binding protein signature PS00880. Since the P-loop occurs in many proteins that bind ATP/GTP, it was the subject of a search. Protein kinases have binding sites for ATP and there is some similarity. The CoA binding domains were sought because CoA influences the time-course of light production by firefly luciferase [39–41].

The domains that can be recognized involve ATP(AMP)-binding sites, regions that interact with ATP, regions that are involved in reactions leading to

Table 4
ProDoms and ProSites found in firefly luciferase

Element	Type	Length, residue	Number proteins	% identity with *Ppe*2	Identified function
ProDom					
3891	Firefly luciferase	30	3	60	Unknown
3894	Firefly luciferase	50	3	52	Unknown
3895	Firefly luciferase	61	3	49	Unknown
3896	Firefly luciferase	51	3	67	Unknown
184	Gramicidin S synthetase	34	45	50	Unknown
185	Gramicidin S synthetase	43	45	42	Unknown
937	4-coumarate-CoA ligase	25	13	45	Unknown
10748	4-coumarate-CoA ligase	27	1	52	Unknown
10747	4-coumarate-CoA ligase	27	1	44	Unknown
10832	2-acylglycerophosphoethanolamine acyltransferase	24	1	42	Unknown
ProSite					
PS00113	Adenylate kinase	12	45	58	Adenylate kinase signature
PS00339	AA tRNA ligase II	10	99	80	Signature
PS00342	Firefly luciferase	3	88	100	C-terminal microbody-directing
PS00455			85		AMP-binding signature
motif 1		12		69	
motif 2		9		67	

% identity with *Ppe*2 is based on the ProDom search or the listed amino-acid residues for the ProSite. ProDom 170 (gramicidin S synthetase) shows two firefly luciferases listed as members of the family, but there appears to be insufficient sequence homology over the whole sequence.

the formation of adenylate intermediates, and function in peptide synthetases. As expected, there are several regions that are found in other firefly luciferases.

4.3. Evolutionary relationships

Wood [42] has reviewed the information on chemical mechanism and evolutionary development of the beetle luciferases. The level of dissimilarity among the beetle luciferases is large; only 27% of the amino-acid sequence is conserved among the cloned beetle luciferases. Wood concluded that the rate of evolution of the luciferases is high relative to other enzymes.

Wood presented a tree diagram showing the relationship of the beetle luciferases with the CoA synthetases [42]. The closest enzyme to the luciferases is 4-coumarate:CoA ligase which has 17% identity in amino-acid sequences. Wood [42] postulates that the firefly luciferases may have evolved from the CoA synthetases. CoA influences the pattern of light production by the luciferases. Without CoA, saturating concentrations of ATP produce a flash of light. CoA prevents the subsequent inhibition of light production and allows a sustained production of light. Wood [40] and Ford et al. [41] found that the -SH group of CoA was required for stimulation of activity. Other nucleotide analogs can produce a steady light production similar to that obtained with CoA [43]. These results have been interpreted as an enhanced turnover of the enzyme mediated by conformational changes. These observations suggest that there may be a vestigial CoA binding site on the luciferases. The crystal structure of Type III chloramphenicol acetyltransferase has been solved at a 1.75 Å resolution by Leslie [44] and the acetyl CoA binding site analyzed [45]. The amino-acid residues interacting with CoA are F-55, Y-56, F-104, S-148, W-152, K-177, Y-178, H-195, and D-199. A similar pattern of residues occurs in *P. pennsylvanica* and other firefly luciferases (the residue number corresponds to the *P. pennsylvanica* sequence and the ratio is the number of times that the residue is found in the aligned position in the 16 luciferases of known sequence): F330, 12/16; Y339, 12/16; F367 6/16 and at position 368, 8/16; W416, 16/16; S419, 11/16 or T, 5/16; K442, 16/16; Y443, 16/16; H460, 12/16; and D465 16/16. These amino acids may constitute the CoA-binding site.

4.4. Crystal structure

The crystal structure of recombinant *P. pyralis* luciferase has been determined at a resolution of 2.0 Å [46]. There are two compact domains — the N-terminal 80% and the C-terminal 20%. The N-terminal domain contains a β-barrel and two β-sheets which are flanked by α-helices. The C-terminal region contains two short antiparallel β-strands and a three-stranded mixed β-sheet, with three helices packed against the side. There is a large cleft between the two parts (domains). Amino acids conserved in the firefly luciferases are in both parts. The P-loop is in a loop connecting antiparallel strands 6 and 7 of β-sheet A. There is a large conformational change when substrates are bound and establishment of a nonpolar environment would insure high quantum yield. Conti et al. [46] suggest that the C-terminal region moves to form a cap. The amino-acid residues that we suggest might be involved in the interaction with CoA are found on both sides of the cleft. There are two peptide sequences that are labeled with radioactive thiourea dioxide, a lysine-reagent [14]. Those two peptides are positioned to be part of the active site — one in the C-terminal domain and the other in the N-terminal domain on faces of the cleft that could meet. The Y-339 and S-419 (*Photuris* numbering) suggested above are predicted to be a part of the *P. pyralis* active site containing (*P. pyralis* sequence numbers) S-198, K-206, Y-340, E-344, E-389, Y-401, S-420, G-421, and D-422 [46].

Baldwin [47], in a review of the firefly luciferase structure that reveals a new protein fold, concludes that 'the mystery remains' as to the molecular mechanism of catalysis. The crystals were obtained without bound substrates or other ligands and the suggested active site is based on those amino-acid residues that are conserved among a superfamily of adenylate-forming enzymes.

Elucidation of the amino-acid residues involved in catalyzing the firefly luciferase reactions awaits the results of specific biochemical modification experiments, additional site-directed mutagenesis studies, and further X-ray crystallographic analysis of the protein structure with bound reactants. The large

conformational changes that occur during catalysis suggest that the dynamics of these changes are important in understanding the mechanism. Firefly luciferase differs from the adenylate-forming enzymes used in the structural comparisons described above because luciferase is a monooxygenase and there must be conserved amino-acid residues for that activity. The tools are rapidly becoming available for a better understanding of this bioluminescent process.

Acknowledgements

This research was supported in part by the Oklahoma Agricultural Experiment Station (Project 1806) and is published with the approval of the director. Thanks to Mses. Ann Williams, Janet Rogers, and Sue Ann Hudiburg of the Sarkeys' Biotechnology Laboratory (Recombinant DNA/Protein Resource Facility) for primer synthesis and automated DNA sequencing. Dr. Don Arnold of the K.C. Emerson Entomology Museum, Oklahoma State University identified firefly species and checked classification schemes. Dr. Ulrich Melcher evaluated the sequencing data and interpretation of the biochemical tree relationships. Dr. Melanie Palmer introduced us to specific molecular biology procedures used with insects and provided materials for practice. Drs. Jerry Devine and Thomas Baldwin provided technical advice on screening clones for luciferase production. The sequences for *Pyrocoelia miyako* and *Hotaria parvula* were provided by Drs. Y. Ohmiya, N. Ohba, H. Toh, and F.I. Tsuji before publication. Dr. Keith Wood provided a correction on our sequence (detected a typographical error that had been entered) and made his unpublished sequences for two *Photuris* luciferases available. Dr. Steve Hartson advised on molecular biology techniques. This paper was read and improved by suggestions from Drs. James Blair, John Cushman, Otis Dermer, Sharon Ford, Robert Matts, Ulrich Melcher, and Keith Wood.

References

[1] Seliger, H.H. and McElroy, W.D. (1960) Arch. Biochem. Biophys. 88, 136–141.
[2] Leach, F.R. (1981) J. Appl. Biochem. 3, 473–517.
[3] Lundin, A. (1982) in Luminescent Assays (Serio, M. and Pazzagli, M., eds.), pp. 29–45, Raven Press, NY.
[4] Kricka, L.J. (1988) Anal. Biochem. 174, 14–21.
[5] Campbell, A.K. (1988) Chemiluminescence: Principles and Applications in Biology and Medicine, Ellis Horwood Ltd., Chichester, 608 pp.
[6] De Wet, J.R., Wood, K.V., Helinski, D.R. and DeLuca, M. (1985) Proc. Natl. Acad. Sci. USA 82, 7870–7873.
[7] de Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) Mol. Cell. Biol. 7, 725–737.
[8] Wood, K.V., Lam, Y.A., Seliger, H.H. and McElroy, W.D. (1989) Science 244, 700–702.
[9] Masuda, T., Tatsumi, M. and Nakano, E. (1989) Gene 77, 265–270.
[10] Tatsumi, H., Kajiyama, N. and Nakano, E. (1992) Biochim. Biophys. Acta 1131, 161–165.
[11] Devine, J.H., Kutuzova, G.D., Green, V.A., Ugarova, N.N. and Baldwin, T.O. (1993) Biochim. Biophys. Acta 1173, 121–132.
[12] Ohmiya, Y., Ohba, N., Toh, H. and Tsuji, F.I. (1995) Photochem. Photobiol. 62, 309–313.
[13] Sala-Newby, G.B., Thomson, C.M. and Campbell, A.K. (1996) Biochem. J. 313, 761–767.
[14] Ye, L. (1994) Firefly Luciferase: Modification and Cloning, Dissertation, Oklahoma State University, Stillwater, OK. Order No. AADAA-19525455, Dissertation Abstracts.
[15] Lloyd, J.B. (1965) Science, 149, 653–654.
[16] Lloyd, J.B. (1984) Fla. Entomol. 67, 368–376.
[17] Strause, L.G., DeLuca, M. and Case, J.F. (1979) J. Insect Physiol. 25, 339–348.
[18] Strause, L.G. and DeLuca, M. (1981) Insect Biochem. 11, 417–422.
[19] Worley, K.C., Wiese, B.A. and Smith, R.F. (1995) Genome Res. 5, 173–184.
[20] Henikoff, S. and Henikoff, J.G. (1994) Genomics 19, 97–107.
[21] Henikoff, S. and Henikoff, J.G. (1991) Nucleic Acids Res. 16, 6565–6572.
[22] Douglas, S.E. (1994) in Computer Analysis of Sequence Data, Part II (Griffin, A.M. and Griffin, H.G., eds.), Methods in Molecular Biology 25, pp. 181–214, Humana Press, Totowa, NJ.
[23] Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1994) Nucleic Acids Res. 22, 3441–3444.
[24] Olson, S.A. (1994) in Computer Analysis of Sequence Data, Part II (Griffin, A.M. and Griffin, H.G., eds.), Methods in Molecular Biology 25, Ch. 17–21, Humana Press, Totowa, NJ.
[25] Bairoch, A. and Boechmann, B. (1991) Nucleic Acids Res. 19, 2247–2249.
[26] Sonnhammer, E.L.L. and Kahn, D. (1994) Protein Sci. 3, 482–492.
[27] Bairoch, A. and Bucher, P. (1994) Nucleic Acids Res. 22, 3583–3589.
[28] Bairoch, A. and Boeckmann, B. (1994) Nucleic Acids Res. 22, 3580–3581.

[29] Herring, P.J. (1987) J. Biolumin. Chemilumin. 1, 147–163.
[30] Jackowski, S., Jackson, P.D. and Rock, C.O. (1994) J. Biol. Chem. 269, 2921–2928.
[31] Toh, H. (1990) Protein Seq. Data Anal. 3, 517–521.
[32] Fulda, M., Heinz, E. and Wolter, F.P. (1994) Mol. Gen. Genet. 242, 241–249.
[33] Fujino, T. and Yamamoto, T. (1992) J. Biochem. 111, 197–203.
[34] Black, P.N., DiRusso, C.C., Metzger, A.K. and Heimert, T.L. (1992) J. Biol. Chem. 267, 25513–25520.
[35] Suzuki, H., Kawarabayasi, Y., Kondo, J., Abe, T., Nishikawa, K., Kimuro, S., Hashimoto, T. and Yamamoto, T. (1990) J. Biol. Chem. 265, 8681–8685.
[36] Pavela-Vrancic, M., Pfeifer, E., van Liempt, H., Schafer, H.-J., Von Doehren, H. and Kleinkauf, H. (1994) Biochemistry 33, 6276–6283.
[37] Maccabe, A.P., van Liempt H., Palissa, H., Unkles, S.E., Riach, M.B.R., Pfeifer, E., Von Doehren, H. and Kinghorn, J.R. (1991) J. Biol. Chem. 266, 12646–12654.
[38] Babbitt, P.C., Kenyon, G.L., Martin, B.M., Charest, H. and Slyvestre, M. (1992) Biochemistry 31, 5594–5604.
[39] Airth, R.L., Rhodes, W.C. amd McElroy, W.D. (1958) Biochim. Biophys. Acta 27, 519–532.
[40] Wood, K.V. (1991) in Bioluminescence and Chemiluminescence: Current Status (Stanley, P.E. and Kricka, L.J., eds.), pp. 543–546.
[41] Ford, S.R., Buck, L.M. and Leach, F.R. (1995) Biochim. Biophys. Acta 1252, 180–184.
[42] Wood, K.V. (1995) Photochem. Photobiol. 62, 662–673.
[43] Ford, S.R., Chenault, K.H., Bunton, L.S., Hampton, G.J., McCarthy, J., Hall, M.S., Pangburn, S.J., Buck, L.M. and Leach, F.R. (1996) J. Biolumin. Chemilumin. 11, 149–167.
[44] Leslie, A.G.W. (1990) J. Mol. Biol. 213, 168–186.
[45] Day, P.J., Shaw, W.V., Gibbs, M.R. and Leslie, A.G.W. (1992) Biochemistry 31, 4198–4205.
[46] Conti, E., Franks, N.P. and Brick, P. (1996) Structure, 4, 287–298.
[47] Baldwin, T.O. (1996) Structure, 4, 223–228.